PEAK
indicators

Getting Real Business Value out of Oracle BI and Oracle Data Mining

Antony Heljula

Brendan Tierney

www.oug.org/ireland
OUG Ireland
Conference 2012
The Convention Centre Dublin | 21st March | 2012

OUG IRELAND
ORACLE USER GROUP

ORACLE APPROVED EDUCATION CENTER

ORACLE PARTNER

Agenda

- Aim of Presentation
- About Oracle Data-Mining
- What you need to know about ODM
- Demonstration of Predictive BI
- Further Considerations
- The BIG Question: Will it work on YOUR data?

- **Aim of Presentation**

Definition

Data Mining is knowledge discovery in data,

discovering non-trivial, hidden and previously unknown information from large collections of data

Aim of Presentation

We've got so much data but we never do any data mining to make sense of what we have!

Once we go live with our dashboards we'll be able to do some complex data-mining!

- How many times have you heard these types of comments?
- Data mining is one of those topics that many people talk about but few actually understand its full potential and what is involved

Aim of Presentation

We've got so much data but we never do any data mining to make sense of what we have!

Once we go live with our dashboards we'll be able to do some complex data-mining!

- This presentation aims to:
  - Describe Data Mining from a "**Business**" perspective
  - Demonstrate the value of integrating Oracle Data Mining and Business Intelligence: "Predictive BI"
  - Explain how any Organization can use and benefit from data mining

- **About Oracle Data-Mining**

Oracle Data Mining

Why do we need it?

- When we analyse data on dashboards, humans are good at spotting simple trends involving a small number of dimensions
  - e.g. Age and Sex

- But humans find it difficult to:
  - Find patterns when you need to analyse large amounts of data over many different dimensions
  - Spot associations or affinities with probabilities for determining a level of confidence

Oracle Data Mining

Why do we need it?

- When we analyse data on dashboards, humans are good at spotting si… involving a… dimensions
  - e.g. Age…
- But human…
  - Find pat… to analyse… data over many different dimensions
  - Spot associations or affinities with probabilities for determining a level of confidence

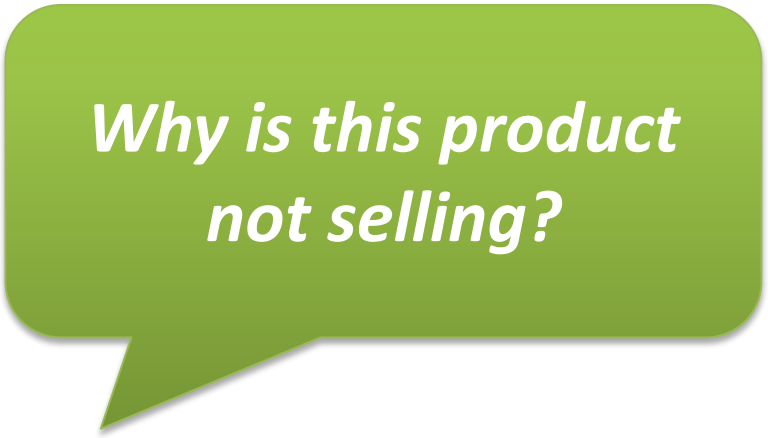
Data mining will help you find hidden and non-trivial patterns of behaviour in your data

Oracle Data Mining

Why do we need it?

- When asked “Business” questions, employees tend to give subjective or biased responses

- Take this example:



Oracle Data Mining

Why do we need it?

We need better demos!

Sales Rep

Too many bugs!

Service Agent

Not enough marketing spend!

Marketeer

Delivery timescales are too tight!

Developer

We don't get enough training!

Consultant

Oracle Data Mining

Why do we need it?

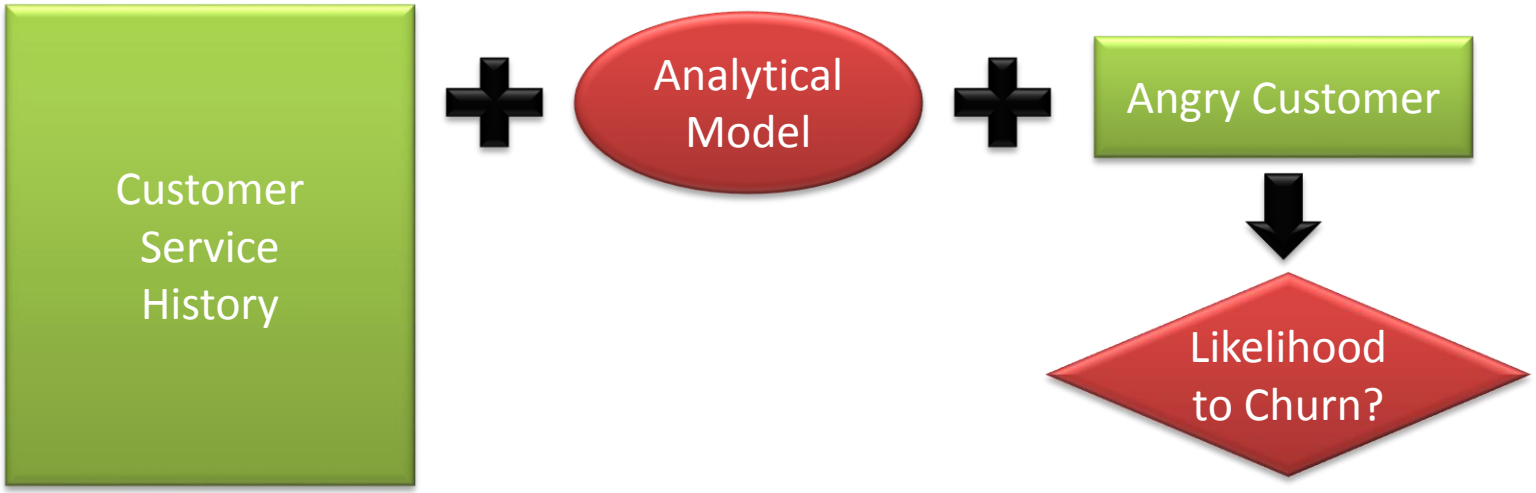
Oracle Data Mining

What does it do?

- Oracle Data Mining (ODM) is a component of the Oracle Database
  - All you data stays in the same place

- With ODM, various algorithms can be applied to your historical data – these algorithms can build complex analytical models to explain your data

- The statistical models can then be used to help answer "Business" questions. For example: Is this angry customer likely to churn?



Oracle Data Mining

Everyone can benefit!

- **Example 1: Human Resources**
  - Build a statistical model on your Employee history:
    - Employee Attributes (grade, office location, organisation etc)
    - Salary increases
    - Bonus payments
    - Termination dates
  - During the next pay review cycle, use the statistical model to predict which employees will leave the company as a result of their pay/bonus awards

Oracle Data Mining

Everyone can benefit!

- **Example 2: Spend Classification**
  - Build a statistical model on your Procurement history to categorise your spending behaviour
  - Use the information to optimise your sourcing decisions
  - For example:
    - Company X realised they bought furniture from 25 different suppliers during the year
    - Negotiating a deal with a single supplier for all office furniture saved £50K each year

Oracle Data Mining

Everyone can benefit!

- **Example 3: Fraudulent or Erroneous Expenses**
  - Build a statistical model on your Expenses history – flagging which expenses were previously found to be either fraudulent or had contained errors
  - When employees submit new expenses, use the statistical model to predict which submissions may need a more detailed review prior to approval

Oracle Data Mining

Everyone can benefit!

- **Example 4: Customer Retention**
  - Build a statistical model on your Customer Service history, flagging customers who have churned to a competitor
  - When customers contact the call centre, use the statistical model to predict the probability of the customer churning

Oracle Data Mining

Everyone can benefit!

- **Example 5: Marketing**
  - Build a statistical model on the history of data generated through your loyalty card scheme
    - Spending behaviour
    - Range of products
    - Combination of products bought together
    - etc
  - For the next marketing campaign, use the statistical model to predict which customers are most likely to be interested in your latest “Pizza + Coke” special deal

Oracle Data Mining

Everyone can benefit!

- **Example 6: Higher Education**
  - Build a statistical model on the history of your student retention
  - When the next intake of students apply for their courses, use the statistical model to predict which students are most likely to withdraw early from their courses

- **What you need to know about ODM**


网易论坛
bbs.163.com

What Do You Need To Know About ODM?

- ……..nothing!
- You just need to know an ODM expert!

ODM and OBI Demonstration

How it happened

Brendan, how can we help universities reduce the number of students who withdraw early from their courses?

Me

What data do you have?

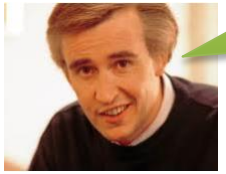
Brendan

We've got a list of all students going back 5 years

Me

Are you able to tell which students withdrew early?

Brendan



Yes, we have a flag to state which ones withdrew early and which ones did not!

Me

ODM and OBI Demonstration

How it happened (continued)

Brendan

Can you give me a data set with 1 record per student with as many attributes as possible, including the "withdraw flag"

Me

OK, I can do that

Brendan

The data set should have 50% withdrawals and 50% non-withdrawals

Me

OK, and then what?

Brendan

I will produce a statistical model that can help you predict which future students are likely to withdraw early

Me

Great!

ODM and OBI Demonstration

Data Set handed to ODM Expert

- 1 record per student, each record with 28 attributes
  - 50% of records with WITHDRAW_FLAG='Y'
  - 50% of records with WITHDRAW_FLAG='N'

ST_PK	WITHDRAW_FLAG	ACADEMIC_CAREER	ACADEMIC_DEGREE	ACADEMIC_ORG	ACADEMIC_PLAN_TYPE	ACADEMIC_PROGRAM	ACADEMIC_YEAR	BIRTH_COUNTRY	STUDENT
24438	Y	UG	FOUND	TECH	FD	HJ39F	0708	GBR	
24442	Y	UG	BA HONS	HEALTH	DEG	B360	0607		
24452	N	UG	PR DIP	EDUC	DEG	UX3AD	0708		
24462	Y	UG	BA HONS	FPL	SP	Y002	0607		
24464	Y	UG	BA HONS	SOCSCI	DEG	L530	0607	NGA	
24490	Y	UG	BSC HONS	SCIENCE	DEG	UC8AE	0607		
24527	N	UG	BSCH PRO	SOCTHER	DEG	B920A	0708	GBR	
24529	N	UG	BSC HONS	SCIENCE	DEG	LL27	0809		
24537	N	UG	LLB HONS	LAW	DEG	M100	0607	GBR	
24565	N	UG	BTEC PD	HEALTH	BTC	UB9AO	0607	GBR	
24577	N	UG	BA HONS	BUSINESS	DEG	UN2AK	0607	ISR	
24607	Y	UG	BA HONS	BUSINESS	DEG	NN13	0809	GBR	
24644	N	UG	BA HONS	BUSINESS	DEG	UN2AK	0708	ISR	
24650	N	UG	BA HONS	FPL	SP	Y002	0809	GBR	
24658	Y	UG	BA HONS	FPL	DEG	Y002	0910	GBR	
24680	N	UG	BED HONS	EDUC	DEG	X100	0809	GBR	
24691	Y	UG	BSC HONS	HEALTH	DEG	UB7AF	0809	GBR	
24698	N	UG	UNI DIP	SCIENCE	SP	UH2AA	0708		
24722	Y	UG	FDSC	TECH	FD	H330	0809	GBR	
24744	Y	UG	HND	BUSINESS	BTC	21NN	0910	GBR	
24755	N	UG	UNI DIP	SHE	DEG	UN8AF	0910		

ODM and OBI Demonstration

Data Mining Model Produced

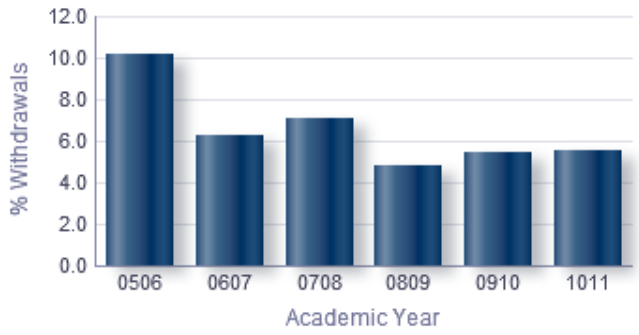
- The statistical model was produced by ODM expert
- Approximately 40% of the attributes were discarded for the analysis
- On completion, we built a simple database view that would return for each new student application:
  - Student Id
  - Withdraw Prediction (Y/N)
  - Prediction Probability

ST_PK	WITHDRAW_PREDICTION	PROBABILITY
46447	N	94.495412844036697
46452	N	66.866840731070498
46599	Y	71.247739602169979
46707	N	66.866840731070498
46776	N	66.866840731070498
46862	Y	78.65168539325843

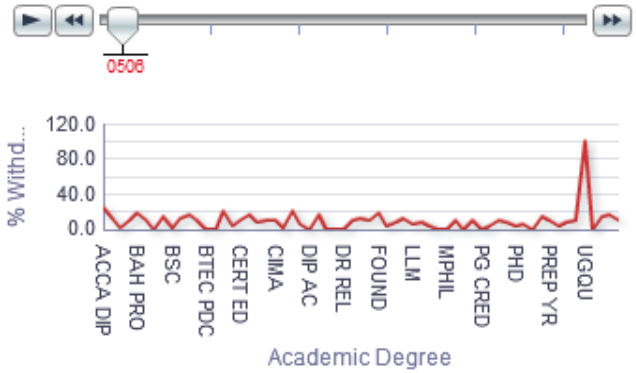
- We could now use these predictions in our dashboards!!

- **Demonstration of Predictive BI**

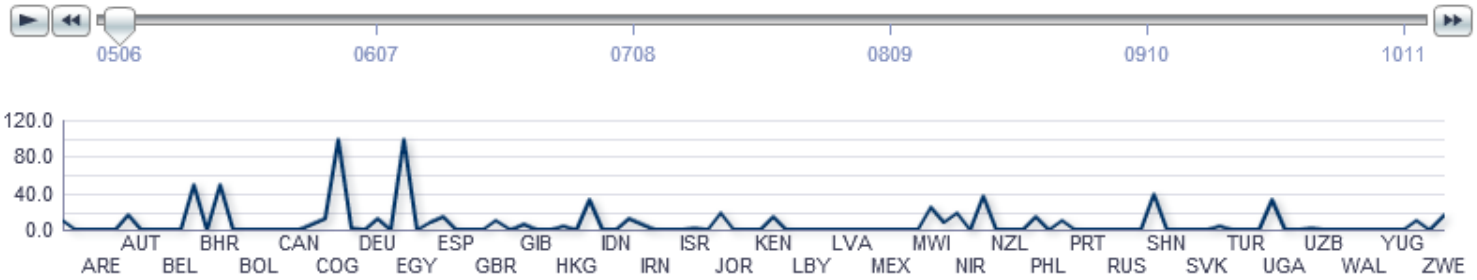
% Withdrawals by Academic Year



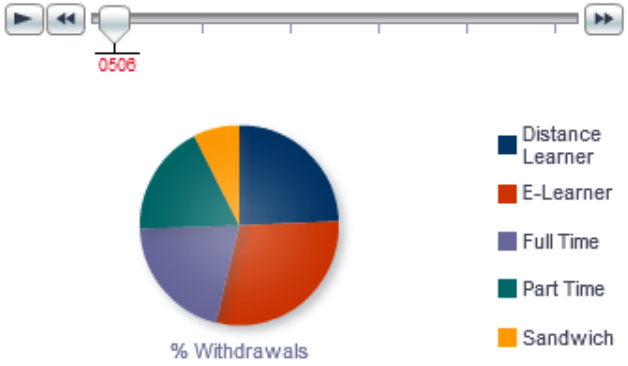
% Withdrawals by Degree



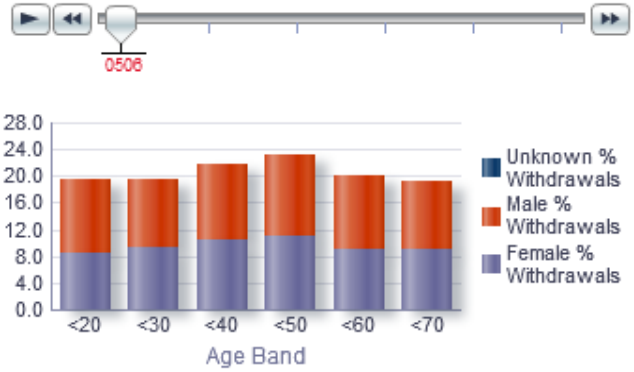
% Withdrawals by Country of Birth



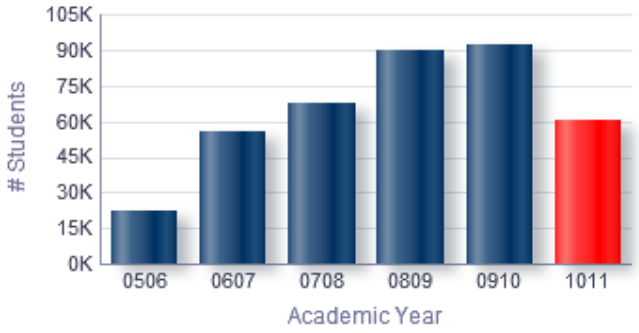
% Withdrawals by Mode of Study



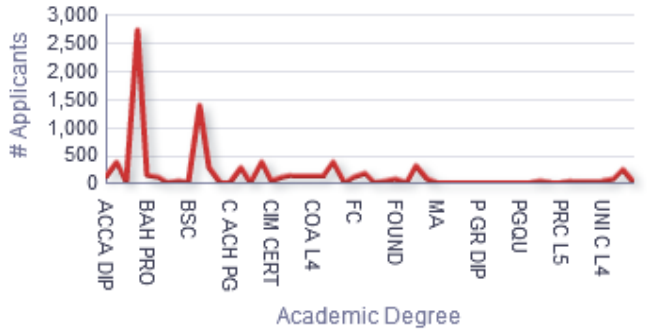
% Withdrawals by Sex and Age



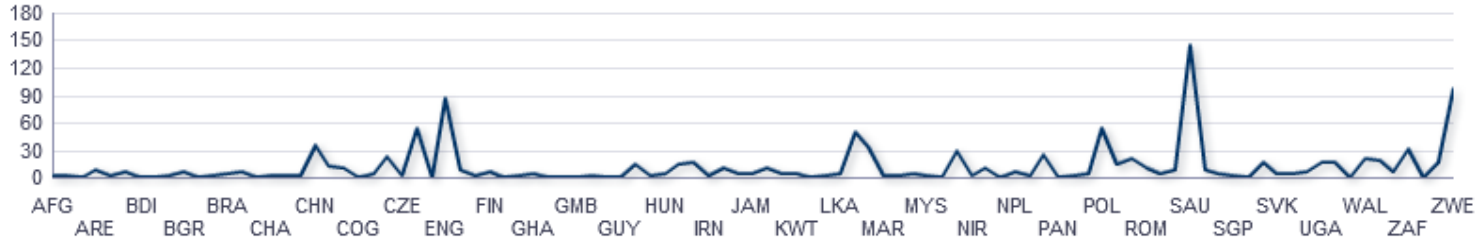
Applicants by Academic Year



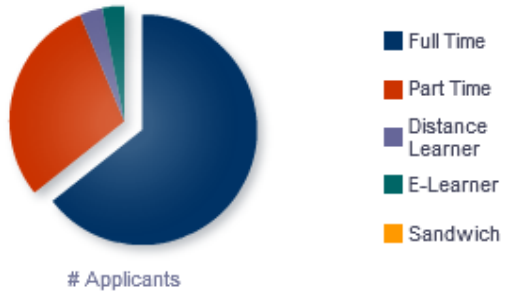
Applicants by Degree



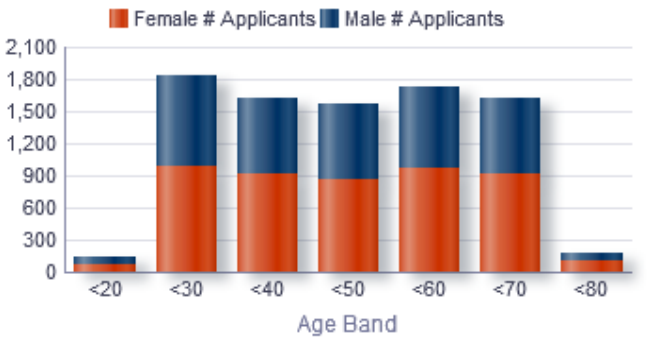
Foreign Applicants by Country of Birth



Applicants by Mode of Study



Applicants by Sex and Age Band



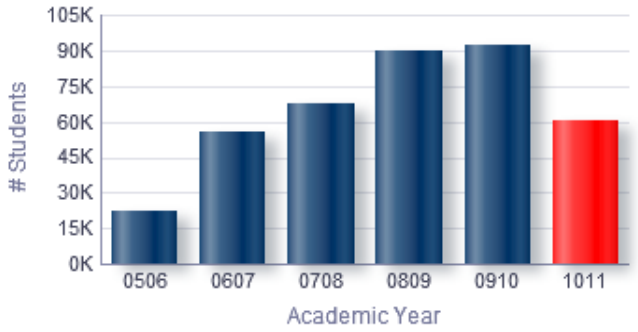
Withdrawal Prediction Filter

Prediction ☐ No ☑ Yes

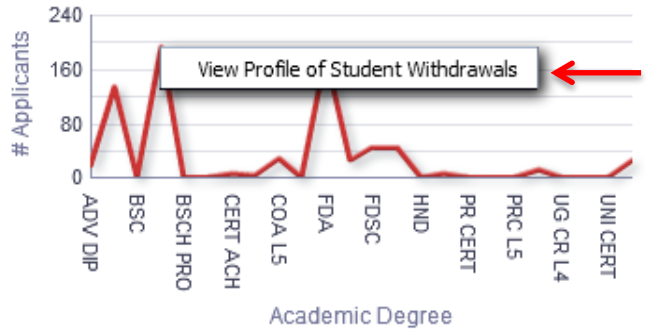
Probability > 70

Apply Reset

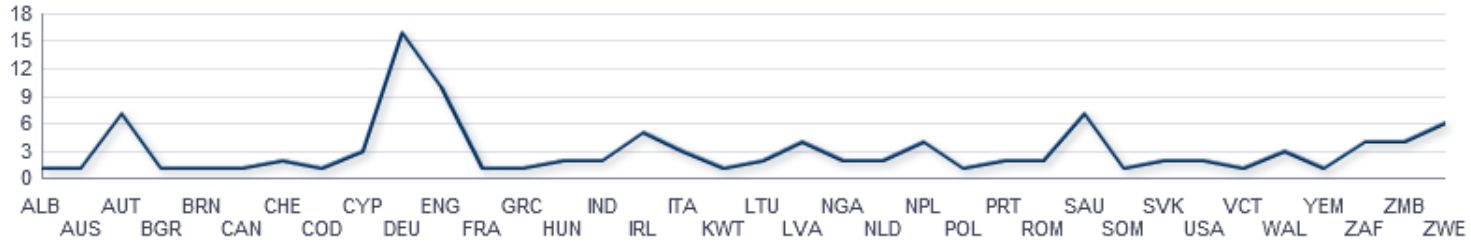
Applicants by Academic Year



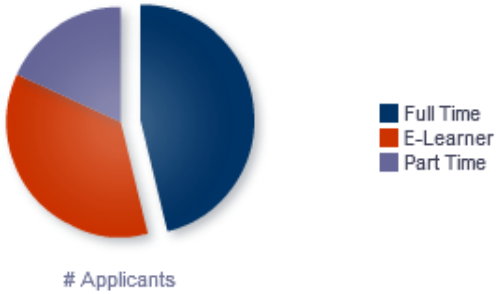
Applicants by Degree



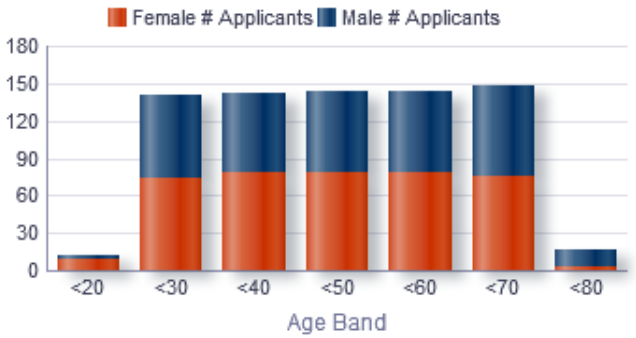
Foreign Applicants by Country of Birth



Applicants by Mode of Study



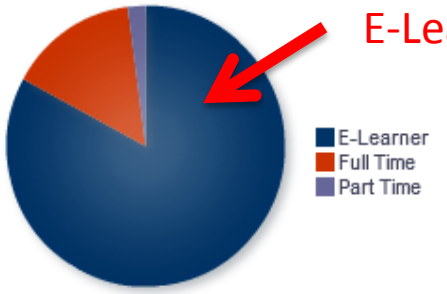
Applicants by Sex and Age Band



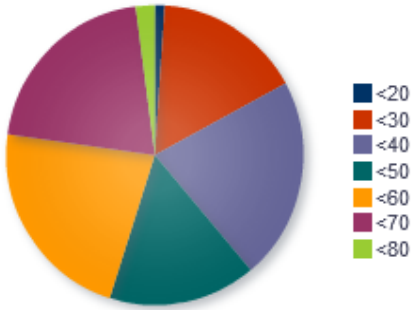
Student Withdrawal Profile

Choose Measure # Withdrawals Predicted

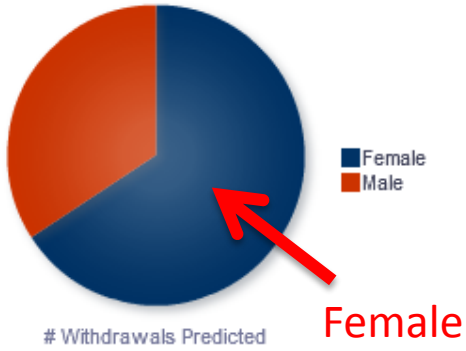
Mode of Study



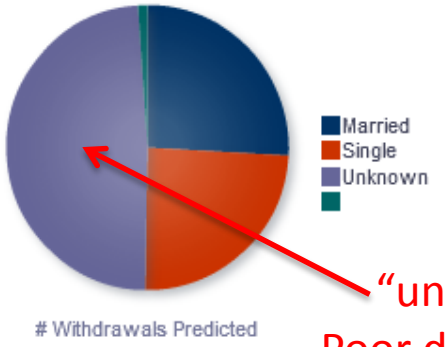
Age Band



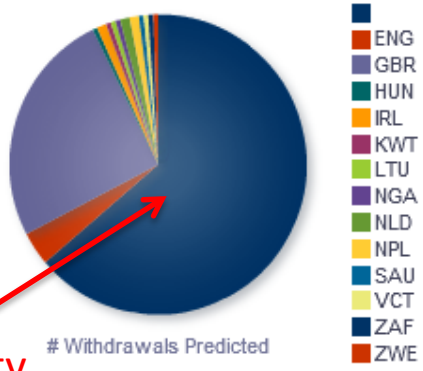
Sex



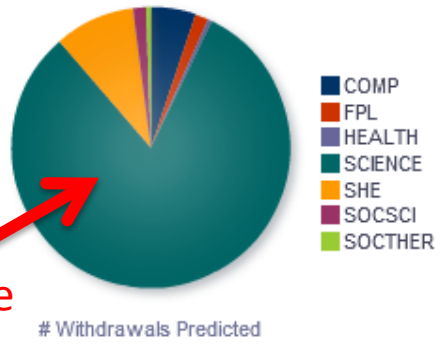
Marital Status



Country of Birth



Academic Org



"unknown"
Poor data quality

Science

Female

Academic Degree is equal to / is in **BSC HONS**
and Withdraw Probability is greater than **70**
and Withdraw Prediction is equal to **Yes**

Student Withdrawal Details

Time run: 3/6/2012 8:48:51 AM

Predictions mostly correct!

Choose View Actual Withdrawals

Withdrawal Prediction: Yes

Student Id	Title	Sex	Marital Status	Country of Birth	Age	Mode of Study	Academic Org	Academic Degree	Academic Program	Withdraw Flag	Withdraw Prediction	Result
3732	Miss	Female	Single		47.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
6951	Mrs	Female	Married		45.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
12022	Miss	Female	Single	GBR	62.00	E-Learner	SCIENCE	BSC HONS	UC8AG	No	Yes	✗
13482	Ms	Female	Unknown		31.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
14436	Miss	Female	Single		66.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
16134	Miss	Female	Single		45.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
19996	Ms	Female	Unknown		48.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
26246	Ms	Female	Unknown		58.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
27615	Ms	Female	Unknown	NLD	58.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
33060	Ms	Female	Unknown	GBR	65.00	E-Learner	SCIENCE	BSC HONS	UC8AG	No	Yes	✗
40254	Mrs	Female	Married		66.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
46163	Mrs	Female	Married		22.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
46862	Mrs	Female	Married		68.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
47353	Mrs	Female	Married		60.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
49969	Miss	Female	Single		22.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
50014	Mrs	Female	Married		24.00	E-Learner	SCIENCE	BSC HONS	UC8AD	No	Yes	✗
58538	Mrs	Female	Married		19.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
61196	Ms	Female	Unknown		54.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
62358	Mrs	Female	Married		58.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
67248	Miss	Female	Single		42.00	E-Learner	SCIENCE	BSC HONS	UC8AG	No	Yes	✗
67678	Miss	Female	Single		62.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
70018	Mrs	Female	Married		65.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
74349	Mrs	Female	Married		39.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
79960	Mrs	Female	Married	GBR	40.00	E-Learner	SCIENCE	BSC HONS	UC8AG	No	Yes	✗
83726	Ms	Female	Unknown		25.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
88221	Ms	Female	Unknown		20.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
93338	Mrs	Female	Married		57.00	E-Learner	SCIENCE	BSC HONS	UC8AG	No	Yes	✗
95117	Mrs	Female	Married		37.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
96737	Mrs	Female	Married		38.00	E-Learner	SCIENCE	BSC HONS	UC8AG	Yes	Yes	✓
97989	Mrs	Female	Married		31.00	E-Learner	SCIENCE	BSC HONS	UC8AG	No	Yes	✗
99825	Miss	Female	Single		57.00	E-Learner	SCIENCE	BSC HONS	UC8AG	No	Yes	✗

- **Further Considerations**

Further Considerations

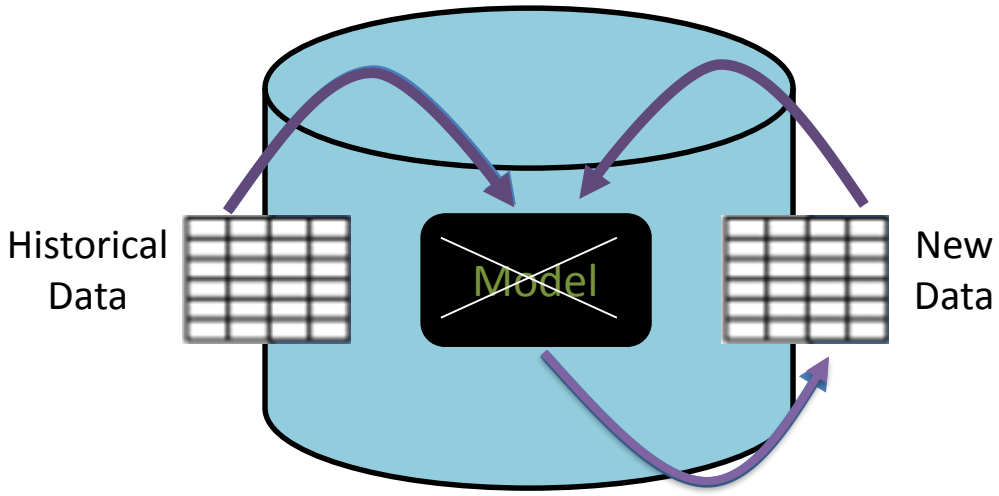
Which ODM Models to Use?

- For this “Higher Education” example, 4 different models were attempted before the right one was found
- How do you know when you have found the right model?
  - With ODM you can partition the data set so that, for example, 80% can be used for building the models and the remaining 20% can be used for testing it afterwards
  - When ODM has finished analysing your data, it tests the resultant statistical model against the bit of data that was not processed
  - You can then keep refining your models until you have a satisfactory amount of “LIFT” (an indicator stating how good or accurate your model is)

Further Considerations

Refining Models

- Do I have to do this all the time?

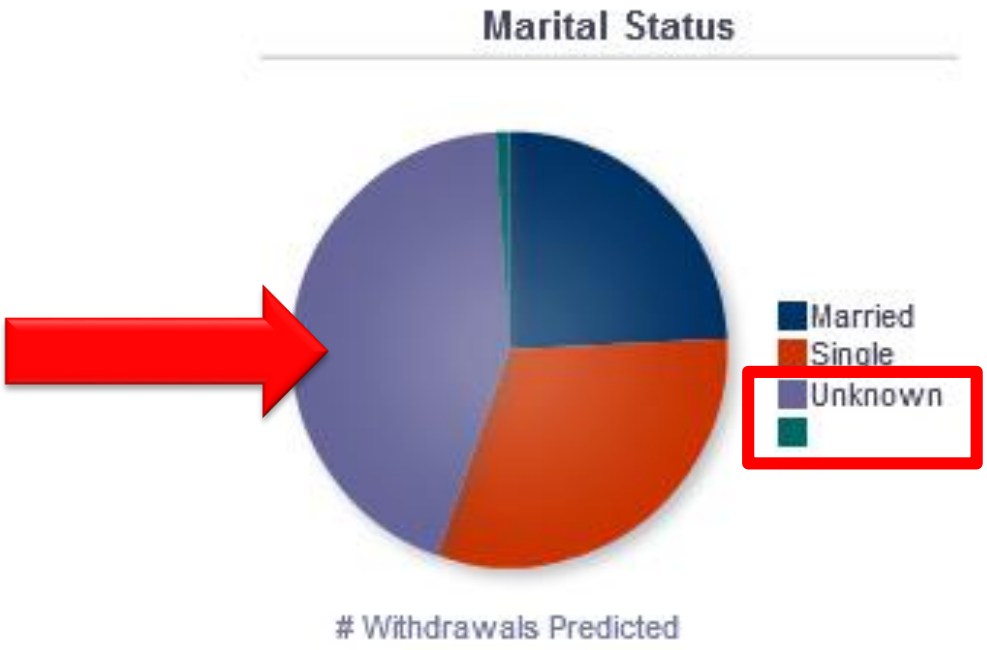


- Answer: Yes and No

Further Considerations

Data Quality

- Data quality is key
- Data cleansing could be a significant part of the whole data mining project
- The amount of “Unknowns” should be reduced to an absolute minimum



- **The BIG Question:**
  Will it work on YOUR data?

The BIG Question!

- **Will it work on YOUR data?**
  - Yes! But…..
  - ODM is ideally suited to a company with a mature (ish) BI environment and good data quality
  - The only real way to prove the effectiveness of ODM is to perform a pilot on your own actual data
  - You should expect a pilot to take approx 3-6 weeks in duration

- **Questions?**

PEAK
indicators

Helping Your Business Intelligence Journey